J. theor. Biol. (2000) **203,** 229–248
doi:10.1006/jtbi.2000.1073, available online at http://www.idealibrary.com on IDEAL®

Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective

CHRISTOPHE H. SCHILLING*, DAVID LETSCHER†‡ AND BERNHARD Ø. PALSSON*§

**Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093-0412, U.S.A. and † Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, U.S.A.*

(*Received on* 27 *December* 1999, *Accepted in revised form on* 25 *February* 2000)

Cellular metabolism is most often described and interpreted in terms of the biochemical reactions that make up the metabolic network. Genomics is providing near complete information regarding the genes/gene products participating in cellular metabolism for a growing number of organisms. As the true functional units of metabolic systems are its pathways, the time has arrived to define metabolic pathways in the context of whole-cell metabolism for the analysis of the structural design and capabilities of the metabolic network. In this study, we present the theoretical foundations for the identification of the unique set of systemically independent biochemical pathways, termed extreme pathways, based on system stochiometry and limited thermodynamics. These pathways represent the edges of the steady-state flux cone derived from convex analysis, and they can be used to represent any flux distribution achievable by the metabolic network. An algorithm is presented to determine the set of extreme pathways for a system of any complexity and a classification scheme is introduced for the characterization of these pathways. The property of systemic independence is discussed along with its implications for issues related to metabolic regulation and the evolution of cellular metabolic networks. The underlying pathway structure that is determined from the set of extreme pathways now provides us with the ability to analyse, interpret, and perhaps predict metabolic function from a pathway-based perspective in addition to the traditional reaction-based perspective. The algorithm and classification scheme developed can be used to describe the pathway structure in annotated genomes to explore the capabilities of an organism.

Introduction

Metabolism is broadly defined as the complex of physical and chemical processes involved in the maintenance of life. It is comprised of a vast repertoire of enzymatic reactions and transport processes used to convert thousands of organic compounds into the various molecules necessary to support cellular life. Metabolic objectives are achieved through a sophisticated control scheme that efficiently distributes and processes metabolic resources throughout the cell's metabolic network. Before we can understand the regulatory logic that the cell chooses to implement, we need to ask what the cell is actually attempting to regulate from a systems-based perspective. The collection of reactions and hence pathways that a metabolic network possesses determines the architecture and topology of the network. To harness the production capabilities inherent in the network, the cell must find a way to control the system and the pathways that determine these capabilities.

‡Present address: Department of Mathematics, Oklahoma State University, Stillwater, OK 74078, U.S.A.
§Author to whom correspondence should be addressed.

0022–5193/00/070229 + 20 $35.00/0

The obvious functional unit in metabolic networks is the actual enzyme or gene product executing a particular chemical reaction or facilitating a transport process. Control of metabolism involves regulation of these individual reactions at various levels from enzymatic activation/inhibition down to transcriptional regulation at the genetic level. By physically controlling the functional units involved in the metabolic network the cell is ultimately controlling its metabolic pathways in a switchboard-like fashion, directing the distribution and processing of metabolites throughout its extensive map of pathways. Thus, in seeking to comprehend the regulatory logic implemented by the cell to control the network it is imperative to understand how the cell is capable of meeting its metabolic objectives through the analysis of its metabolic pathways.

Decades of metabolic research and the advent of rapid genome sequencing technologies and algorithms have placed us on the brink of having a complete metabolic parts catalogue for several organisms. Once a genome has been fully sequenced and annotated, the entire metabolic map representing all the metabolic reactions taking place in the cell can be constructed, and extensive on-line databases have collected this information for a number of organisms (Kanehisa, 1997; Karp *et al*., 1998; Selkov *et al*., 1998). This information provides us with the contents of the metabolic system in a given organism. Concurrent advances in the area of cDNA microarrays (Schena *et al*., 1998; Brown & Botstein, 1999) and DNA chip technology (Hoheisel, 1997; Ramsay, 1988; Lipshutz *et al*., 1999) have provided the capability of studying the expression patterns and utilization of the metabolic genotype under various environmental conditions. Additionally, computational models of the entire metabolic network are now being developed to analyse, interpret, and predict the genotype–phenotype relationship for fully sequenced organisms (e.g. Schilling *et al*., 1999a).

All of these methods for the analysis of metabolic networks provide clear and insightful information regarding the activity of metabolic reaction networks from an individual reaction-based perspective. From the microarray data we can observe which genes have been up- or down-regulated under changing conditions, and from the computational approaches it is predicted which fluxes increase or decrease under simulated conditions. It is now an opportune time to provide a new perspective for the analysis of metabolic networks, namely that of a pathway-oriented interpretation. Thus, we seek to translate information on the activity of individual reactions in a metabolic network into distinct metabolic pathways. Ideally, we would like to find a set of pathways that are unique for a particular network, which correspondingly can provide a unique representation of all the possible functional modes or flux distributions that the network can achieve. With such a set of pathways defined for a given network of reactions, we could use these pathways to provide the pathway-based perspective that is needed to understand how metabolic networks are operated and controlled.

Currently, there exist two fundamentally different approaches to the definition of metabolic pathways: (1) qualitative identification based on historical groups of reactions in a database setting (Karp *et al*., 1999), and (2) rigorous quantitative and systemic definitions based on mathematical principals such as linear algebra and convex analysis (for a recent review see Schilling *et al*., 1999b).

While much information can be derived from a qualitative analysis of the pathways in a metabolic network, comprehensive approaches for the quantitative definition of metabolic pathways can be used to assess the precise metabolic capabilities and performance of cellular metabolic networks under a broad range of environmental and genetic challenges. This will be illustrated in the following companion paper (Schilling & Palsson, 2000) in which the theoretical approaches outlined below are applied to examine the production capabilities and functional characteristics of *Haemophilus influenzae*, containing the first fully sequenced genome of a free-living organism (Fleischmann *et al*., 1995).

In this paper, we present the detailed theoretical foundations of a newly developed approach for the study of metabolic pathway analysis that is based on convex analysis, which has been previously used to study pathways in both chemical and metabolic reaction networks (Clarke, 1988; Schuster *et al*., 1999). Through the application of convex analysis, the unique set of extreme pathways that are systemically independent can be

calculated for any metabolic system and then classified. These pathways can then be used to interpret metabolic functioning and potentially metabolic control. A section devoted to a comparison between existing approaches for pathway analysis based on principles of convexity is also included at the end. Mathematical proofs and instructive examples are provided to illustrate many of the concepts discussed in an effort to make the theory intelligible to anyone with an interest in either biology or mathematics. The application of these theoretical concepts to a fully sequenced and annotated genome will follow (Schilling & Palsson, 2000).

Metabolic Systems and Descriptions

A cellular metabolic reaction network is a collection of enzymatic reactions and transport processes that serve to replenish and drain the relative amounts of certain metabolites. A system boundary can be drawn around all these types of physically occurring reactions, which constitute internal fluxes operating inside the network. The system is closed to the passage of certain metabolites while others are allowed to enter and/or exit the system based on external sources and/or sinks which are operating on the network as a whole. The existence of an external source/sink on a metabolite necessitates the introduction of an exchange flux, which serves to allow a metabolite to enter or exit the theoretical system boundary. These fluxes are not physical biochemical conversions or transport processes like those of internal fluxes, but can be thought of as representing the inputs and outputs to the system. These fluxes can also be referred to as pseudoreactions (Clarke, 1980), and could also represent diffusive exchange with a buffered external reservoir. Demands on a metabolite for further processing or incorporation into cellular biomass would also create an exchange flux on an internal cellular metabolite. (As a consequence, a metabolite located inside the cell is considered distinct from the same metabolite located in the extracellular space.) Furthermore, when considering a whole-cell metabolic network the system boundary is drawn around the entire cell and transport reactions become internal fluxes (similar to the concept of a free-body diagram in mechanical engineering for determining force balances, or shell balances in transport phenomena).

All internal fluxes are denoted by v_i for $i \in [1, n_I]$ where n_I is the number of internal fluxes. All exchange fluxes are denoted by b_i, for $i \in [1, n_E]$ where n_E is the total number of exchange fluxes. Limited thermodynamic information can be used to determine if a chemical reaction can proceed in the forward and reverse directions or it is irreversible thus physically constraining the direction of the reaction. All internal reactions that are considered to be capable of operating in a reversible fashion are considered as two fluxes occurring in opposite directions, therefore constraining all internal fluxes to be nonnegative. This convention is used purely for mathematical purposes and does not influence the biological interpretation of metabolic function in any way. There can only be one exchange flux per metabolite, whose activity subsequently represents the net production and consumption of the metabolite by the system. Thus, n_E can never exceed the number of metabolites in the system (m). The activity of these exchange fluxes is considered to be positive if the metabolite is exiting or being produced by the system, and negative if the metabolite is entering or being consumed by the system. For all metabolites in which a source or sink may be present the exchange flux can operate in a bidirectional manner and is therefore unconstrained.

The analysis of a metabolic system should begin with a study of its structural characteristics or invariant properties, those depending neither on the state of the environment nor on the internal state of the system, but only on its structure (Reder, 1988). The stoichiometry of a biochemical reaction network is the primary invariant property that describes the architecture and topology of the network. Stoichiometry refers to the molar ratios in which substrates are converted into products in a chemical reaction (e.g. glucokinase converts one mole of glucose and one mole of ATP into one mole of glucose-6-phosphate and one mole of ADP). These ratios remain constant under changing reaction conditions, which may serve to alter the kinetic parameters and rate of reaction as a function of time.

Dynamic mass balances can be written around every metabolite in the system taking the form of the following equation in matrix notation where **x** denotes the concentration vector of all the metabolites, **S** is the stochiometric matrix and **v** is the flux vector describing the activity of all the internal and exchange fluxes:

$$\frac{d\mathbf{x}}{d\mathbf{t}} = \mathbf{S} \cdot \mathbf{v}. \tag{1}$$

The stoichiometric matrix **S** is an $m \times n$ matrix where m corresponds to the number of metabolites and n is the total number of fluxes taking place in the network ($n = n_I + n_E$). The S_{ij} element of the stochiometric matrix corresponds to the stoichiometric coefficient of the reactant i in the reaction denoted by j, and v_j is the flux through this metabolic reaction. Thus, for the enzyme glucokinase, $S_{glucose,glucokinase} = -1$. The vector **v** then refers to the relative fluxes through the reactions in the metabolic network. In these differential equations, enzyme kinetics enters the equation when **v** is represented as a rate law, typically a function of the concentration vector and a number of kinetic parameters. Equation (1) then essentially describes the change in metabolite concentration as a function of time to be equal to the differences in the sum of all fluxes that serve to produce the metabolite and those which consume the metabolite.

The pathway structure we seek to determine should also be an invariant property of the network along with stoichiometry. Thus, it is reasonable to eliminate the time derivative from eqn (1) by imposing a steady-state condition. Under steady-state conditions the time derivative in eqn (1) can be relaxed to zero and a resulting set of linear homogeneous equations [eqn (2)] is created from which it is possible to calculate flux values. This system of equations is typically underdetermined for metabolic systems as the number of reactions typically exceeds the number of participating metabolites, and as a result a corresponding null space (Nul **S**) can be described (Lay, 1997):

$$\mathbf{0} = \mathbf{S} \cdot \mathbf{v}. \tag{2}$$

The null space corresponds to the set of all solutions (**v**) for eqn (2). It has been previously shown that a set of basis vectors can be selected to describe the null space of eqn 2, where each basis vector corresponds to a steady-state biochemical pathway (Fell, 1993; Schilling & Palsson, 1998). However, to completely describe the system we need to include the constraints on internal and exchange fluxes. The constraints on internal fluxes are rather straightforward as all fluxes must be non-negative yielding:

$$v_i \geqslant 0, \forall i. \tag{3}$$

The constraints on the exchange fluxes depend on the status of a determined source or sink on the metabolite, or similarly on the input and output status of the metabolite. These constraints can be expressed as shown in eqn (4) where α_j and β_j are either zero or negative and positive infinity, respectively, based on the direction of the exchange flux. Under the existence of a source (input) only α_j is set to negative infinity and β_j is set to zero, whereas if only a sink (output) exists on the metabolite α_j is set to zero and β_j is set to positive infinity. If both a source and a sink are present for the metabolite then the exchange flux is bidirectional with α_j set to negative infinity and β_j set to positive infinity leaving the exchange flux unconstrained.

$$\alpha_j \leqslant b_j \leqslant \beta_j \tag{4}$$

In general, we can make the distinction between currency metabolites (i.e. cofactors such as ATP, NADH) involved in energy and redox levels and the rest of the metabolites in the network (primary metabolites). For a pathway analysis the exchange fluxes for these currency metabolites are typically unconstrained. Under conditions in which the system is to be closed for the exchange of these metabolites the corresponding exchange fluxes are constrained to zero. This distinction between metabolites will be of assistance in classifying metabolic pathways. Additionally, we typically structure the stoichiometric matrix so that the first series of columns represent the internal fluxes and the remaining columns represent the exchange fluxes with the primary exchange fluxes followed by the currency exchange fluxes. In constructing the stoichiometric matrix in this way, the vector **v** is composed

first of entries $v_I - v_{n_I}$ followed by $b_I - b_{n_E}$. This construction has the advantage that the overall systemic balance equation for a given pathway or flux distribution is simply obtained from the values of the exchange fluxes constituting the lower portion of the flux vector (i.e. see Fig. 2 and the discussion below).

These general concepts can be illustrated by a specific example. Figure 1(a) depicts a biochemical reaction network consisting of five metabolites and four internal reactions interconverting these metabolites. Two of these reactions are reversible creating a forward flux and a reverse flux. Together there are a total of six internal fluxes. Only four of the five metabolites have sources or sinks and are subsequently allowed to cross over the system boundary creating four exchange fluxes that are all unconstrained. Figure 1(b) shows the complete mathematical translation of the system into eqns (2)–(4). It should be noted that such a translation and the determination of a genome-specific stoichiometric matrix representing an entire cellular metabolic network can be directly determined from an organisms metabolic genotype and biochemical information based on the reactions and processes associated with each of the respective gene products of the genotype (Edwards & Palsson, 1998; Schilling *et al*., 1999a).

The Steady-state Flux Cone and Metabolic Capabilities

Together eqns (2)–(4) describe a metabolic system under steady-state conditions as a system of linear equalities/inequalities. This description

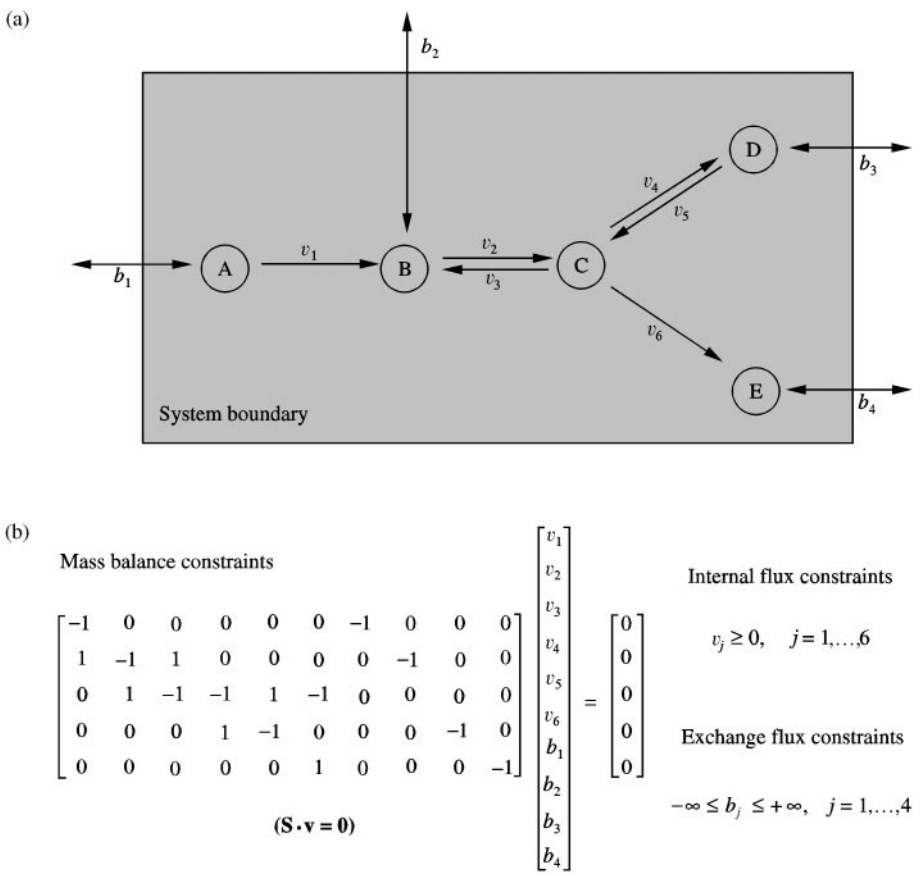


FIG. 1. (a) Chemical reaction network consisting of the five metabolites (A–E) and four internal reactions (two reversible) creating six internal fluxes (v) along with four allowed exchange fluxes (b) all indicated by the arrows. (b) The translation of the network into its mathematical representation. The stoichiometric matrix (**S**) is expressed in terms of the series of linear homogeneous equations derived from the conservation of mass in eqn (2). All internal fluxes are constrained to be nonnegative as in eqn (3) and the exchange fluxes are all unconstrained and described as in eqn (4).

captures the constraints that are placed on the network by the stoichiometry and thermodynamics of the reactions, as well as the constraints on input and output of metabolites from the system, which typically can be experimentally determined. The presence of linear inequalities limits the use of traditional concepts of linear algebra, and necessitates the use of convex analysis (Rockafellar, 1970), which is capable of treating systems of linear inequalities. The set of solutions to any system of linear inequalities is a convex set, and for our particular system this is also true as any linear equality can be written as two inequalities. This convex solution corresponds geometrically to a convex polyhedral cone in n-dimensional space (R^n) emanating from the origin for all metabolic systems modeled as described herein. We refer to this convex cone generally as the flux space and more specifically as the steady-state flux cone (C). Within this flux cone lie all of the possible steady-state solutions and hence the flux distributions under which the system can operate. Since every solution or operating mode of the system is contained within the flux space, it logically follows that the entire flux space represents the capabilities of the given metabolic network. Thus, the flux space clearly defines what a network can and cannot do; what building blocks can be manufactured; how efficient the energy extraction and conversion of carbohydrates into ATP can be for a given substrate; where are the critical links in the network; and so on. The answer to these basic questions and many others related to the structural and functional capabilities of the network are found within the flux cone.

As the answers to our questions lie within the flux cone we must then develop a way to describe and interpret any location within this space. In other words, we must now find the best way to navigate through this solution space. We are mainly interested in determining the characteristics of this space and interpreting it from an overall metabolic perspective. This objective can be achieved by either interpreting the functioning of the network from the traditional reaction-based perspective as described by the flux vector $\mathbf{v}$ or from a pathway-oriented perspective. We now discuss the development of such a pathway-oriented perspective.

Convex Analysis and Metabolic Pathways

The study of convex polyhedral cones, which forms the underlying mathematical structure for metabolic pathway analysis, has several conceptual similarities with linear algebra. For convex cones, one studies extreme rays (or generating vectors) that correspond to edges of the cone being half-lines emanating from the origin. These extreme rays are said to generate the cone and cannot be decomposed into a non-trivial convex combination of any other vectors residing in the flux cone. For this reason they are referred to as being conically or systematically independent. This notion of a minimal generating set, which is properly referred to as forming the conical hull of the cone, roughly corresponds to the concept of a basis in linear algebra. However, this generating set is typically unique, providing a clear advantage to using convex analysis rather than just studying the underlying linear algebra coming from the subspace determined by the null space of $\mathbf{S}$. In fact, the set of conditions for describing convex cones is nearly identical to the conditions used to define a vector space in linear algebra with the exception that all scalars must be non-negative for convex cones (Hadley, 1961).

Here in the context of metabolic systems we will use the term extreme pathways to denote the extreme rays of a polyhedral cone as each ray corresponds to a particular pathway or active set of fluxes which satisfies the steady-state mass balance constraints and inequalities posed in eqns (2)–(4). Extreme pathways will be denoted by the vector $\mathbf{p}_i$ and the total number of extreme pathways needed to generate the flux cone for a system will be denoted by k. Every point within this cone (C) can be written as a non-negative linear combination of the extreme pathways as

$$C = \left\{ \mathbf{v} : \mathbf{v} = \sum_{i=1}^{k} w_i \mathbf{p}_i, \quad w_i \geqslant 0 \ \forall i \right\}. \qquad (5)$$

Thus, the set of extreme pathways is analogous to a basis/coordinate system that can be used to describe a position in space. These pathways are said to conically span or generate the set of all pathways as any pathway or distribution of fluxes can be written as a non-negative linear

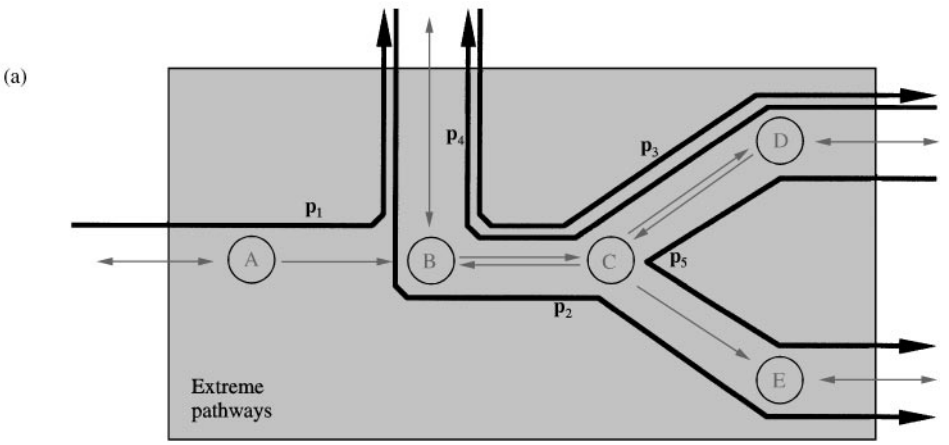
combination of the $\mathbf{p}_i$'s. (Notice the similarity with the concept of a spanning or generating set in linear algebra.) The pathway vector **w** corresponds to the coordinate vector relative to the set of extreme pathways. It provides the weight given to each pathway in a particular flux distribution (**v**). If we consider the matrix **P** whose columns are composed of the set of extreme pathways, this matrix conceptually transforms the flux vector **v** into the pathway vector **w** providing a pathway-based perspective of the functioning of the network as opposed to a reaction-based perspective.

The flux cone in R^n can be geometrically thought of as the intersection between the null space of eqn (2) and the vector space described by the inequalities of eqns (3) and (4). The vector space described by eqn (3) is the positive orthant in n_I-dimensional space, while the vector space described by eqn (4) is a region of n_E-dimensional space. Thus, the flux space can be described as the following convex subset of R^n:

$$C = (R_+^{n_I} \times R^{n_E}) \cap (\mathrm{Nul}\,\mathbf{S}), \tag{6}$$

In other words, the flux cone contains all points of the null space whose coordinates are non-negative, with the exception of the exchange fluxes that are constrained to be negative or those that are unconstrained.

To illustrate these concepts we return again to the system described in Fig. 1. Figure 2(a)



(b)

$$\mathbf{P} = \begin{array}{c} \begin{matrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 & \mathbf{p}_5 & \mathbf{p}_6 & \mathbf{p}_7 \end{matrix} \\ \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ \hdashline -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}\right] \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ \hline b_1 \\ b_2 \\ b_3 \\ b_4 \end{array} \end{array}$$

(c) Flux distribution:

$$\mathbf{v}^{\mathrm{T}} = [4\ 2\ 0\ 1\ 0\ 1 - 4\ 2\ 1\ \ 1]$$

Subset pathways of **v**:

$$\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$$

Null space dimensions of $\mathbf{S}_{mod}$:

$$d(\mathbf{S}_{mod}) = n - r(\mathbf{S}_{mod}) = 8 - 5 = 3$$

Unique decomposition of **v**:

$$\mathbf{w}^{\mathrm{T}} = [4\ 1\ 1\ 0\ 0\ 0\ 0]$$

or

$$\mathbf{v} = (4)\cdot\mathbf{p}_1 + (1)\cdot\mathbf{p}_2 + (1)\cdot\mathbf{p}_3$$

FIG. 2. (a) The set of extreme pathways for the network described in Fig. 1. Five of the seven extreme pathways are illustrated corresponding to the edges of the flux cone of admissible steady-state flux vectors. Only Type I pathways are illustrated. The two additional pathways not depicted (Type III) correspond to the cycling of the two reversible reactions (v_2/v_3 *and* v_4/v_5). (b) All seven pathways are presented as columns in the pathway matrix **P**. The net systemic balance equations for each pathway are obtained from the value of the exchange fluxes below the dashed line. (c) An example of a flux distribution (**v**) is given to illustrate the test for a unique decomposition of a flux vector. The number of subset pathways of **v** equals the dimensions of the null space for the modified stoichiometric matrix creating a unique decomposition.

graphically illustrates five of the seven extreme pathways for the reaction network described in Fig. 1, while Fig. 2(b) provides the matrix **P** whose seven columns are the extreme pathways. The algorithm that is used to calculate the extreme pathways from eqns (2)–(4) is presented in Appendix B. Using these seven pathways all possible flux distributions can be decomposed as shown in eqn (5). Later we will discuss a pathway classification scheme.

Unique Extreme Pathways

The first issue that is of concern with any set of pathways used to describe a metabolic network is the uniqueness of the set. The set of generating vectors for a polyhedral cone is certainly unique when the cone lies in the positive orthant as is the case when all fluxes are constrained to be positive, but what happens when some of these constraints are relaxed on the exchange fluxes and the flux cone no longer lies entirely in the positive orthant? In this case, the set of extreme pathways is still unique and this is stated in the following theorem.

Theorem. *A convex flux cone determined by eqns* (2)–(4) *has a set of systematically independent generating vectors. Furthermore, these generating vectors* (*extremal rays*) *are unique up to multiplication by a positive scalar. These generating vectors will be called extreme pathways.*

The proof of this theorem is given in Appendix A. A simple way to understand and test for the uniqueness property under any set of flux conventions is to set all unidirectional fluxes (constrained to be non-negative) equal to zero and search for a non-trivial solution to eqn (2). If a solution exists then this confirms the existence of a real line in the solution space described by eqn (6) resulting in a loss of uniqueness. If we consider all fluxes to be constrained as non-negative then there will obviously be no solution other than the trivial solution when these fluxes are set to zero, indicating that the set of edges of the convex flux cone is unique. Uniqueness can be lost when both reversible internal and exchange fluxes are described as bidirectional fluxes with no constraints. As an example, if we lumped the forward and reverse fluxes v_2/v_3 and v_4/v_5 together into two unconstrained reversible fluxes, the set of extreme pathways would no longer be unique as a non-trivial solution would exist when the unidirectional fluxes v_1 and v_6 are removed, thus confirming the presence of a real line in the solution space. To preserve uniqueness we can easily adopt the convention of decomposing the reversible internal fluxes into two unidirectional fluxes constrained to be nonnegative while allowing exchange fluxes to be free of constraints under conditions where a source and sink are accounted for. If all unidirectional fluxes are set to zero there will be no solution to eqn (2) other than the trivial solution as all internal fluxes will be set to zero and the exchange fluxes must then equal zero since only one exchange flux is present per metabolite. Therefore, the set of extreme pathways is observed to be unique under the adopted conventions.

The conclusion is that under the conventions described herein a unique set of pathways can be described which provide the simplest unique view of the pathway structure of the network offering an additional perspective from which to interpret metabolic function.

Classification of Extreme Pathways

In determining the set of extreme pathways for a system there are, in general, two distinct classes of pathways. There are pathways for which the coefficients of the exchange fluxes are all equal to zero, and there are pathways in which non-zero values exist for a set of exchange fluxes. Furthermore, if a distinction is made between the currency metabolites and the primary metabolites of the system, a third class of pathways can be delineated as those pathways for which all of the exchange fluxes for the primary metabolites equal zero while non-zero values exist for the exchange fluxes of some of the currency metabolites.

The major pathways that are of functional interest are those for which the exchange fluxes of the primary metabolites are active. These pathways are the major contributors to the decomposition of almost any steady-state flux distribution, and will be classified as Type I pathways. Type II pathways will denote those in which only the

exchange fluxes for the currency metabolites are active. These pathways correspond to true futile cycles existing within the network which serve to dissipate energy or reductive power. Perhaps the best example of this class would be the well-known futile cycle that exists in glycolysis between the activity of phosphofructokinase and fructose-1,6-bisphosphatase which dissipates energy by converting ATP into ADP and inorganic phosphate in equal molar ratios. It should be noted that the term futile might be misleading in the rare case in which an internal-energy-producing or redox-producing cycle may exist.

Pathways in which all of the exchange fluxes are inactive correspond to internal cycles within the network that have no net overall effect on the functional capabilities of the network. We will classify these pathways as Type III pathways. Usually, these pathways will correspond to the cycling of two fluxes resulting from the decomposition of reversible reactions into two unidirectional fluxes for the forward and reverse reactions. On occasion a cycle comprised of multiple (> 2) active internal fluxes may exist, but it will again have no impact on the capabilities of the network. While these pathways may not appear to provide a path through the system they still constitute an edge of the flux cone. When interested in providing a decomposition of a flux distribution, these fluxes will virtually never appear in the decomposition as they do not affect the overall productive capabilities. However, these pathways cannot be completely ignored, as there may be dynamic consequences that are yet to be investigated such as their ability to dynamically regulate metabolite or metabolite pool concentrations. A matrix representation of these three different classifications of pathways is shown in Fig. 3. This classification scheme can be extended to classify any pathway or flux distribution in addition to the extreme pathways of a network.

For the reaction of Fig. 1 the first five columns of the matrix **P** in Fig. 2(b) correspond to Type I pathways and are all illustrated in Fig. 2(a) while the last two columns of **P** correspond to Type III pathways, as all of the exchange fluxes are inactive. As all of the metabolites were considered primary, there does not exist any Type II pathway. With these pathways defined and classified how can they be used to analyse and interpret systemic functions and flux distributions?

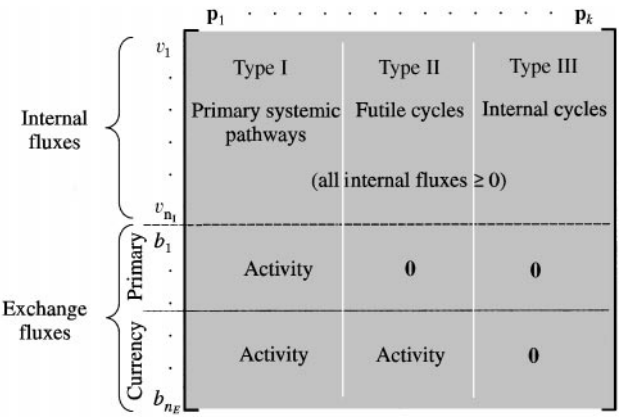


FIG. 3. A matrix perspective for the classification scheme of extreme pathways. The matrix depicted is the pathway matrix **P** where each column represents an extreme pathway. [For an example see Fig. 2(b).] For Type I pathways the only requirement is for one of the primary exchange fluxes to be active. In Type II pathways only the currency exchange fluxes can be active, and in Type III pathways none of the exchange fluxes are active.

Unique Representation of Steady States

Every flux distribution, **v**, can be written as a non-negative linear combination of the extreme pathways, as shown in eqn (5). In linear algebra such a decomposition as a sum of basis vectors is unique even though the basis itself is non-unique. However, some flux distributions can be written as a sum of extreme pathways in many ways. Therefore, the decomposition of a steady-state flux vector (**v**) into the corresponding extreme pathways (**p**) is not necessarily unique. Only a basis for a solution space guarantees a unique representation of every point in the solution space. For the set of extreme pathways to form a basis the number of pathways must equal the dimensions of the null space. The dimensions of the null space depend on the number of free variables in the original set of linear equations forming eqn (2), which is referred to as the rank of **S**, (r). The relation termed the Rank Theorem gives the dimensions of the null space (d):

$$d(\mathbf{S}) = \dim(\mathrm{Nul}\,\mathbf{S}) = n - r. \tag{7}$$

Thus, for a full rank matrix the dimensions of the null space will be equal to the difference between

the number of fluxes and metabolites ($d = n - m$), as r will equal m. If the number of extreme pathways exceeds the dimension of the null space ($k > d$), then the pathways do not uniquely describe every point in the solution cone. In this case, one can select a subset of the extreme pathways which are linearly independent and equal in number to the dimensions of the null space as a basis if desired. This corresponds to the edges of a simplex of the cone. However, this selection is again non-unique as often there are numerous simplexes in which a solution may lie. The edges of the cone therefore can be thought of as providing a limited set of vectors from which to select for the construction of a basis if desired.

If we again consider the reaction network discussed in Fig. 1, the dimension of the null space is 5 and the number of edges of the flux cone is 7 ($d = 5$, $k = 7$). Therefore, the entire flux cone cannot be uniquely decomposed into the extreme pathways. To have the number of pathways equal to the dimension of the null space is uncommon in larger networks due to the high degree of interconnectivity amongst metabolites and reactions.

Even though in most cases the entire cone cannot be uniquely described by the set of extreme pathways, there are certain regions of the solution cone in which a solution is uniquely described by the pathways. To determine if a decomposition is unique it is necessary to first determine the number of extreme pathways that are subsets of the particular flux distribution of interest, **v**. For an extreme pathway to be a subset of a flux distribution it must not contain an active internal flux that is inactive in the flux distribution. Additionally, the extreme pathway must not have an active exchange flux that is inactive in the flux distribution and is also constrained to be either positive or negative. As an example, consider the flux distribution ($\mathbf{v}^T = [4, 2, 0, 1, 0, 1, -4, 2, 1, 1]$) shown in Fig. 2(c). The only inactive fluxes are v_3 and v_5. Thus any pathway that has either of these internal fluxes active is not a subset pathway. This eliminates extreme pathway $\mathbf{p}_4$, $\mathbf{p}_5$, $\mathbf{p}_6$, and $\mathbf{p}_7$, leaving only the first three extreme pathways as subsets of **v**. The next step involves calculating the dimensions of the null space for the modified stoichiometric matrix ($\mathbf{S}_{mod}$) wherein all columns corresponding to inactive internal and exchange fluxes in **v** are removed. In our example, this would eliminate the columns corresponding to v_3 and v_5. The dimension of the corresponding null space is 3, which is equal to the number of subset pathways, thus ensuring a unique decomposition. If the number of extreme that are subsets of **v** is greater than the dimensions of the null space, $d(\mathbf{S}_{mod})$, then the decomposition is non-unique. Otherwise the decomposition is unique.

Systemic Independence

A set of pathways $\{\mathbf{p}_1, \ldots, \mathbf{p}_k\}$ is said to be conically or systemically independent if no pathway can be written as a non-trivial non-negative linear combination of the other pathways. Notice that the only difference between this definition and linear independence is that the coefficients of the linear combination cannot be negative. An important implication of this definition is that it is possible for a set of pathways to be systemically independent while simultaneously being a linearly dependent set. When investigating the functional aspects of a metabolic system the property of systemic independence should take priority over linear independence as it is a unique property of any system and its structural capabilities.

The set of extreme pathways for a system is not linearly independent since typically the number of pathways that form the edges of the steady-state flux cone is greater than the dimensions of the null space. However, the set of extreme pathways is systemically independent by definition of being the edges of the flux cone. The term genetic independence was first introduced to describe a distinction between sets of metabolic pathways which were linearly dependent but defined as independent genotypes (Seressiotis & Bailey, 1988). The idea reflected the notion that certain sets of pathways share a characteristic of systemic independence even though they are linearly dependent. The use of the term genetically independent may potentially be misleading, as there is not necessarily a one-to-one correspondence between reactions and gene products. There are a number of examples of gene products capable of carrying out a number of related but distinct biochemical transformations.

What are the implications of pathways being systemically independent? From a control point of view, it can be envisioned that if a cell could control the activity of each one of the extreme pathways then it would be capable of reaching each point within the flux cone. In other words, it would be maximizing its metabolic capabilities. This is most likely not the case physiologically as a distinct control switch is most likely not devoted to each pathway even in a non-obvious manner. The control scheme that the cell chooses to implement is much more complicated, as the regulation of certain pathways will be coupled and in fact there will be certain pathways that are always inactive or even unregulated due to the regulatory scheme implemented. Certain combinations of pathways may be unfeasible. The challenging task that now confronts us is to understand how the pathways of a system are regulated.

The real question we are seeking to answer is can we gain insight on the regulatory logic implemented by the cell by focusing on its pathway structure. Is the set of extreme pathways what the cell is truly aimed at controlling and regulating? And if so, how are they regulated? Rather than predicting a control scheme that could be used for a metabolic network it will be perhaps most beneficial to look at the actual known regulatory scheme of certain well-understood networks and make some rational connections to the theoretical pathway structure which can be determined. It seems logical to think that these pathways are the ultimate objective of cellular regulation. Thus, while regulation is occurring at the protein and genetic levels through enzymatic activation/inhibition, and transcriptional control, etc., the ultimate cellular function of these regulatory mechanisms is the control of metabolic pathways in an indirect and non-intuitive manner.

Discussion

The definition and conceptualization of biochemical pathways in the context of a whole cell has emerged as an important issue now that genomics is leading to the complete definition of genotypes in an increasing number of organisms. Here we have introduced a theoretical framework for the identification of the simplest unique set of biochemical pathways for a metabolic system, based on methods of convex analysis and the laws of material conservation. The unique extreme pathways are the systemically independent pathways of the network and can be used to analyse and interpret the functional capabilities and flux distributions of cellular metabolism from a pathway-based perspective. The application of this theoretical approach to the problem of analysing a complete cellular metabolic network derived from its metabolic genotype will be discussed in a paper to follow (Schilling & Palsson, 2000). The detailed approach discussed herein represents a continuing improvement in theoretical strategies for the study of metabolic pathways.

The theory behind most recent work on pathway analysis stems from linear algebra and more specifically convex analysis, two related branches of mathematics (for a review see Schilling *et al.*, 1999b). The first natural approach to study metabolic networks is through the use of linear algebra to explore the null space of the series of linear homogeneous equations resulting from the conservation of mass as in eqn (2). As previously mentioned, the null space can be spanned by a set of linearly independent basis vectors which correspond to biochemical pathways that can be used to interpret the functional characteristics of the system (Fell, 1993; Schilling & Palsson, 1998). While basis vectors provide a unique representation of every solution to the system, the set of basis vectors that span the null space is non-unique making their use much less effective. To overcome this obstacle of non-uniqueness we have turned to the mathematics associated with convex spaces.

Convex analysis was first applied to inorganic chemical systems in the detailed theory of stoichiometric network analysis (SNA) (Clarke, 1980, 1981, 1988). This theory was developed for the mathematical analysis of stability in complex reaction networks. SNA utilizes convex analysis to determine a set of "extreme currents" which serve as a framework for a coordinate transformation used to determine the stability of the network. These "extreme currents" also correspond to edges of a flux cone; however, in SNA all fluxes are constrained to be non-negative. Thus,

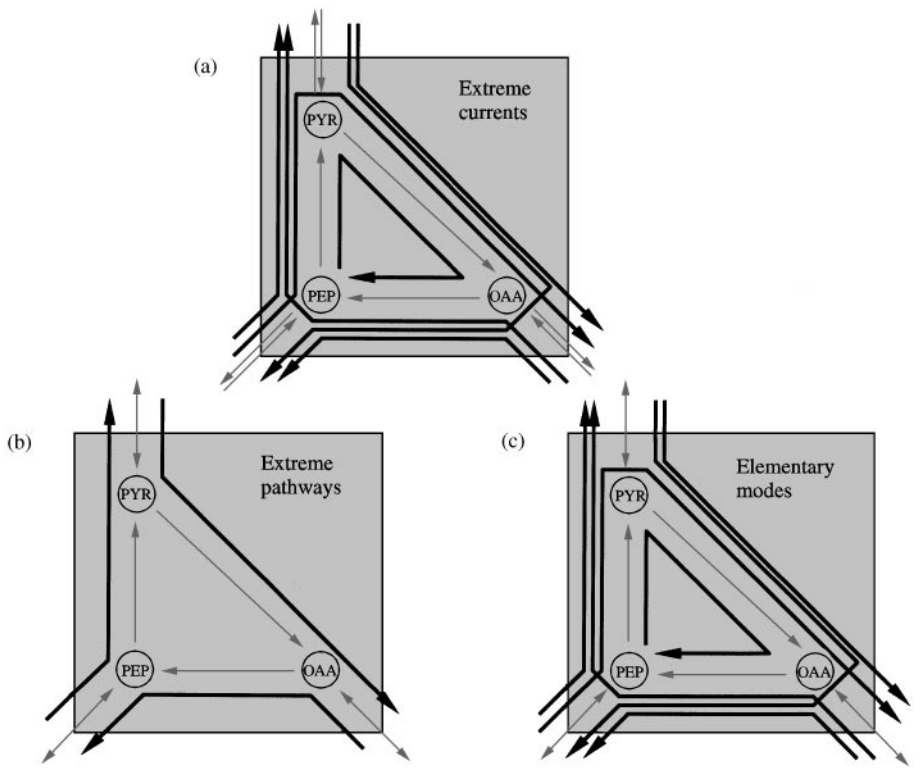


FIG. 4. Scheme of the PEP/PYR/OAA cycle. Abbreviations: PYR, pyruvate; PEP, phosphoenolpyruvate; OAA, oxaloacetate. (a) Seven extreme currents of pathways of the system considering all fluxes to be unidirectional including exchange fluxes. Omitted are the three pathways each of which corresponds to the cycling of the input and output exchange fluxes for each of the metabolites. (b) Three extreme pathways for the system when considering exchange fluxes to be unconstrained or bidirectional. (c) Seven elementary modes for the system when considering exchange fluxes to be unconstrained or bidirectional.

exchange fluxes are decomposed into an independent input flux and an output flux. The effects of this can be seen in Fig. 4 for the simple reaction system representing the pyruvate/phosphoenolpyruvate/oxaloacetate cycle. When all fluxes are considered to be non-negative there are ten edges to the flux cone, seven of which are illustrated in Fig. 4(a). From a systemic point of view a number of the pathways shown are combinations of other pathways, however mathematically this is not a true statement for the exclusive reason that the pseudoreactions (exchange fluxes) are decomposed into forward and reverse reactions. In the approach described herein these pseudoreactions are effectively merged into a bidirectional unconstrained flux creating a flux cone that contains only three edges/pathways shown in Fig. 4(b).

Allowing exchange fluxes to become bidirectional has the general effect of reducing the dimensions of the null space (d) in addition to reducing the number of extreme pathways or edges of the flux cone (k) to an even greater extent. The quantity ($k - d$) therefore decreases. This serves to reduce the degrees of freedom that can be used to select a set of pathways to form a basis. So, while in most cases we do not arrive at a unique basis we can limit the number of different bases to choose from significantly with this approach, bringing us as close to a unique basis as possible if one is seeking to use such a basis to interpret metabolic function.

This appears to be the more reasonable approach, as it provides a much cleaner view of the pathway structure. Additionally, the pseudoreactions should not be influencing the pathway structure to such an extent. The ramifications of changing the conventions used to describe a reaction network on the subsequent operations of SNA and pathway stability are a topic of future investigation. SNA has been most notably used

to establish a categorization of oscillatory reactions (Schreiber *et al*., 1996), the stoichiometric connectivity of chemical species essential for the oscillations (Eiswirth *et al*., 1991), and the quantitative determination of extreme currents active at stationary states in a model of the chloride–iodide reaction at various external constraints (Strasser *et al*., 1993).

The most recent application of convex analysis for the study of metabolic pathways has been in the development of the concepts of elementary flux modes of a system (Schuster & Hilgetag, 1994; Schuster *et al*., 1999). The concepts of elementary modes were most notably utilized to guide the development and engineering of an *Escherichia coli* strain that successfully channeled carbohydrates down the pathways for aromatic amino acid biosynthesis at theoretical yields (Liao *et al*., 1996). Elementary modes have been defined as the minimal set of enzymes that could operate at steady state with all irreversible reactions proceeding in the appropriate directions. Rule-based algorithms utilizing the principles of convex analysis have been previously described for determining the set of elementary modes for situations in which reversible reactions are modeled as bidirectional fluxes (Schuster *et al*., 1996). While the conventions for describing the metabolic system are slightly different from those discussed here, the set of elementary modes has been shown to be unique; however, in the presence of reversible reactions there are often more elementary modes than are needed to span the flux cone. At times this situation will make the decomposition of a flux vector into elementary modes non-unique.

The elementary modes of the system described in Fig. 4(a) with exchange fluxes decomposed are identical to the edges of the flux cone due to the absence of any bidirectional fluxes, and correspond exactly to the extreme currents or pathways of the network as determined using SNA. Using the conventions adopted herein for describing the network as shown in Fig. 4(b), the number of elementary modes is greater than the number of extreme pathways indicating that a number of these elementary modes lie within the interior of the flux cone generated by the extreme pathways and are positive combinations of the extreme pathways. The elementary modes for the network are shown in Fig. 4(c). These modes all correspond to the same pattern of pathways shown in Fig. 4(a). The details of determining the elementary modes for this cycle have been previously discussed (Schuster & Hilgetag, 1994). Here the dimension of the null space is 3 and the number of extreme pathways is 3; however the number of elementary modes is 7. Three of these modes are the same as the extreme pathways while the remaining four lie on a face of the flux cone or within the interior of the flux cone.

Figure 5 illustrates the elementary modes for the reaction network shown in Fig. 1. For both cases the dimension of the null space is 5. All of the extreme pathways are identical to the elementary modes with the two additional elementary modes being the pathways leading from metabolites A to D and A to E. These two additional modes are both positive linear combinations of two extreme pathways and thus lie on the interior of the cone or on a face of the cone in this case. This creates a situation in which there is a redundancy in the pathway structure resulting often in a non-unique decomposition of a steady-state flux distribution.

With respect to SNA and the study of elementary modes the work presented here can be seen as a sort of hybrid approach that builds upon many of the concepts of these two similar approaches looking forward to the use of pathway analysis to study metabolic function in addition

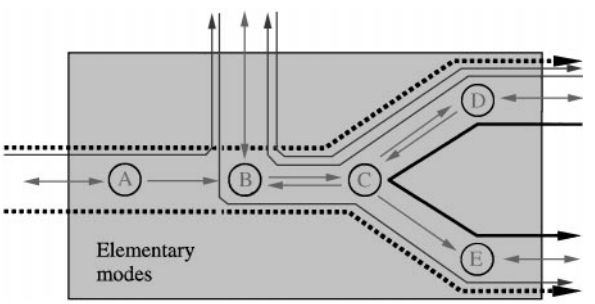


FIG. 5. Illustration of the set of elementary modes corresponding to the reaction network of Fig. 1. Seven elementary modes are shown. All of the elementary modes correspond to the extreme pathways of the network with the exception of the two pathways indicated by the dashed arrows. The upper mode corresponds to the combination of extreme pathways $\mathbf{p}_1$ and $\mathbf{p}_3$ while the lower mode is the combination of extreme pathways $\mathbf{p}_1$ and $\mathbf{p}_2$. Two additional modes exist corresponding to the cycling of the two reversible reactions (v_2/v_3 and v_4/v_5).

to structural aspects of metabolic networks. In studying the extreme pathways we arrive as close as possible to a unique pathway structure that can be used to uniquely describe certain points within the flux cone. These pathways represent the smallest unique set of pathways that can possibly be generated to accurately interpret the functional aspects of a metabolic network. These characteristics are achieved due to the concept of systemic independence and the inability to decompose these pathways into any non-negative combination of other pathways present within the flux cone. Thus, we arrive at the underlying pathway structure and topology of metabolic networks. It should once again be noted that there is no dynamic or regulatory information accounted for in the network description and calculation performed herein, which are based primarily on the structure and topology of the network.

Conclusion

Individual metabolic reactions and fluxes within the cell are the unit of chemical function, while individual extreme pathways may be considered as the unit of systemic function and perhaps cellular function. Here we have discussed how metabolic systems may be described mathematically, and how to determine and classify the unique set of these extreme pathways that correspond to the edges of the flux cone. The underlying pathway structure that is determined from the set of extreme pathways now provides us with the ability to analyse, interpret, and predict metabolic function from a pathway-based perspective.

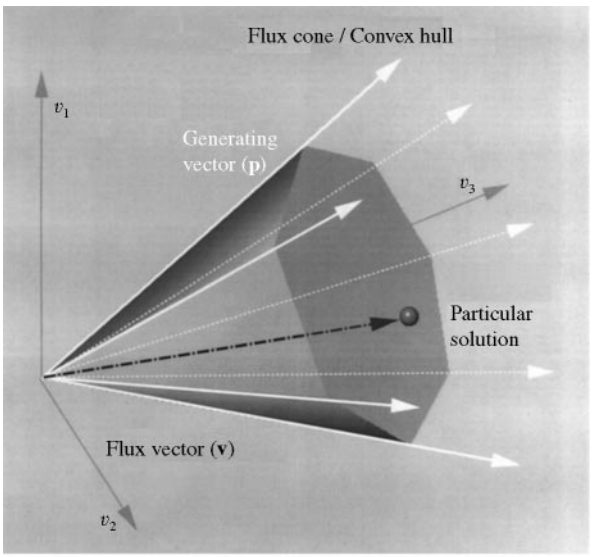


FIG. 6. Defining the metabolic genotype and phenotype in the context of convex analysis. A geometric depiction of the flux cone in three dimensions where the entire unbounded flux cone corresponds to the theoretical capabilities of a metabolic genotype. Each edge of the cone indicated by the white arrows corresponds to an extreme pathway. A particular solution or metabolic phenotype is indicated by the filled point that is described by a flux vector (**v**) lying on the interior of the flux cone or convex hull. This flux vector can be decomposed into extreme pathways as in eqn (5).

Figure 6 provides a geometric interpretation of the flux cone with every point described by eqn (5). Together the set of extreme pathways describes the full capabilities of the metabolic network in the simplest form possible, as these pathways are systemically independent and irreducible. From the biological point of view, the entire flux cone associated with the reactions comprising cellular metabolism corresponds to the capabilities of an organism's metabolic network and hence the capabilities of its metabolic genotype. Each one of the generating vectors corresponds to an extreme pathway that the cell could theoretically control to reach every point in the flux cone. Each particular point within this flux cone corresponds to a different flux distribution representing a particular metabolic phenotype. The actual flux vector describing that point can be thought of as a positive combination of these extreme pathways. So one may think of these pathways as theoretically being "switched" off and on to varying degrees to reach a particular metabolic phenotype. How this control is actually implemented at the enzymatic and genetic levels is the next question. To add capabilities to its metabolic network an organism must

acquire enough reactions to add another extreme pathway. Thus, acquiring a gene whose gene product performs a specific reaction(s) may not add function to the cell if other supporting reactions are not acquired to create a new extreme pathway. Clearly, this has a number of interesting implications regarding the evolution of cellular metabolism throughout the kingdoms of life.

Perhaps one of the greatest scientific challenges of the next century will be to understand the principles and control schemes that underlie integrated multi-geneic functions and the genotype–phenotype relationship (Palsson, 1997). To meet the challenge, we must first understand the systematic objectives of cellular functions and how these objectives are achieved on the cellular level in addition to the molecular level. Then we can proceed with the investigation of the regulatory aspects of these networks and various analysis and simulation methods (McAdams & Arkin, 1998; Tomita *et al*., 1999). With genomics and bioinformatics now providing much of the "hardware" within the cell, we now begin the search for the "software" or the operating systems that govern coordinated cellular function.

The authors wish to acknowledge Jeremy Edwards, John Ross, and Stefan Schuster for their comments and contributions following many discussions on pathway analysis and related topics. Portions of this work are supported by grants from the National Institute of Health (GM57089) and the National Science Foundation (9873384, 9814092). Christophe Schilling is a recipient of a graduate fellowship from the Whitaker Foundation and would like to thank the Whitaker Foundation for their support through graduate fellowships in bioengineering.

REFERENCES

Brown, P. O. & Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nature Genet*. **21**(1 Suppl), 33–37.

Clarke, B. L. (1980). Stability of complex reaction networks. *Adv. Chem. Phys*. **43,** 1–215.

Clarke, B. L. (1981). Complete set of steady states for the general stoichiometric dynamical system. *J. Chem. Phys*. **75,** 4970–4979.

Clarke, B. L. (1988). Stoichiometric network analysis. *Cell Biophys*. **12,** 237–253.

Edwards, J. S. & Palsson, B. O. (1988). How will bioinformatics influence metabolic engineering. Biotechnol. Bioeng. **58**, 162–169.

Eiswirth, M., Freund, A. & Ross, J. (1991). Mechanistic classification of chemical oscillators and the role of species. *Adv. Chem. Phys*. **80,** 127.

Fell, D. A. (1993). The analysis of flux in substrate cycles. In: *Modern Trends in Biothermokinetics* (Schuster, S., Rigoulet, M., Ouhabi, R. & Mazat, J.-P., eds), pp. 97–101. New York: Plenum.

Fleischmann, R. D. & Adams, M. D. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenza* Rd. *Science* **269,** 496–512.

Hadley, G. (1961). *Linear Algebra*. Reading, MA: Addison-Wesley Publishing Co.

Hoheisel, J. D. (1997). Oligomer-chip technology. *Trends Biotechnol*. **15,** 465–469.

Kanehisa, M. (1997). A database for post-genome analysis. *Trends Genet*. **13,** 375–376.

Karp, P. D., Riley, M., Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A. & Merrick, J. M. (1998). EcoCyc: encyclopedia of *Escherichia coli* genes and metabolism. *Nucl. Acids Res*. **26,** 50–53.

Lay, D. C. (1997). *Linear Algebra and Its Applications*. Reading, MA: Addison–Wesley Longman Inc.

Liao, J.C., Hou, S.-Y. & Cho, Y.-P. (1996) Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng*. **52,** 129–140.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet*. **21** (1 Suppl), 20–24.

McAdams, H. H. & Arkin, A. (1998). Simulation of prokaryotic genetic circuits. *Ann. Rev. Biophys. Biomol. Struct*. **27,** 199–224.

Palsson, B. O. (1997). What lies beyond bioinformatics? *Nat. Biotechnol*. **15,** 3–4.

Ramsay, G. (1998). DNA chips: state-of-the-art. *Nat. Biotechnol*. **16,** 40–44.

Reder, C. (1988). Metabolic control theory: a structural approach. *J. theoret. Biol*. **135,** 175–201.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton, NJ: Princeton University Press.

Schena, M., Heller, R. A., Karp, P. D., Riley, M., Paley, S. M., Pellegrini-Toole, A. & Krummenacker, M. (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol*. **16,** 301–306.

Schilling, C. & Palsson, B. (2000). Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. theor. Biol*. **203,** 247–281.

Schilling, C. H., Edwards, J. S. & Palsson, B. O. (1999a). Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog*. **15,** 288–295.

Schilling, C. H. & Palsson, B. O. (1998). The underlying pathway structure of biochemical reaction networks. *Proc. Nat. Acad. Sci. U.S.A*. **95,** 4193–4198.

Schilling, C. H., Schuster, S., Palsson, B. O. & Heinrich, R. (1999b). Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog*. **15,** 296–303.

SCHREIBER, I., HUNG, Y.-F. & ROSS, J. (1996). Categorization of some oscillatory enzymatic reactions. *J. Phys. Chem.* **100,** 8556–8566.
SCHUSTER, R. & SCHUSTER, S. (1993). Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Comput. Appl. Biosci.* **9,** 79–85.
SCHUSTER, S., DANDEKAR, T. & FELL, D. A. (1999). Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* **17,** 53–60.
SCHUSTER, S. & HILGETAG, C. (1994). On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **2,** 165–182.
SCHUSTER, S., HILGETAG, C., WOODS, J. H. & FELL D. A. (1996). Elementary modes of functioning in biochemical networks. In: *Computation in Cellular and Molecular Biological Systems* (Cuthbertson, R., Holcombe, M. & Paton, R., eds), pp. 151–165. London: World Scientific.
SELKOV, E. J., GRECHKIN, Y., MIKHAILOVA, N. & SELKOV, E. (1998). MPW: the metabolic pathways database. *Nucl. Acids Res.* **26,** 43–45.
SERESSIOTIS, A. & BAILEY, J. E. (1988). MPS: an artificially intelligent software system for the analysis and synthesis of metabolic pathways. *Biotechnol. Bioeng.* **31,** 587–602.
STRASSER, P., STEMWEDEL, J. D. & ROSS, J. (1993). Analysis of a mechanism of the chloride–iodide reaction. *J. Phys. Chem.* **97,** 2851–2861.
TOMITA, M., HASHIMOTO, K., SCHENA, M., HELLER, R. A., THERIAULT, T. P., KONRAD, K., LACHENMEIER, E. & DAVIS, R. W. (1999). E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15,** 72–84.

APPENDIX A

Theorem. *A convex flux cone has a set of systemically independent generating vectors. Furthermore, these generating vectors (extremal rays) are unique up to a multiplication by a positive scalar. These generating vectors will be called extreme pathways.*

Proof. First, we must show existence of a systemically independent generating set for a cone and then we will prove uniqueness. To show existence we can simply refer to the algorithm outlined in Appendix B that is justified and shown to provide a set of systemically independent set of generating vectors for a cone. Thus existence is readily shown. To show uniqueness, let $\{\mathbf{p}_1, \ldots, \mathbf{p}_k\}$ be a systemically independent generating set for a cone. Notice that if

$$\mathbf{p}_j = \mathbf{c}' + \mathbf{c}'' \tag{A.1}$$

then both $\mathbf{c}'$ and $\mathbf{c}''$ are positive multiples of $\mathbf{p}_j$. To see why this is true, we can write the two pathways as non-negative linear combinations of the extreme pathways:

$$\mathbf{c}' = \sum_{i=1}^{k} \varsigma_i \mathbf{p}_i \quad \text{and}$$

$$\mathbf{c}'' = \sum_{i=1}^{k} \imath_i \mathbf{p}_i \quad \text{for } \varsigma_i, \imath_i \geqslant 0. \tag{A.2}$$

Thus

$$\mathbf{p}_j = \mathbf{c}' + \mathbf{c}'' = \sum_{i=1}^{k} (\varsigma_i + \imath_i)\mathbf{p}_i. \tag{A.3}$$

Since the $\mathbf{p}_i$ are systemically independent, the only way this can happen is if all the η_i and μ_i are equal to zero except for η_j and μ_j. This implies that both $\mathbf{c}'$ and $\mathbf{c}''$ are multiples of $\mathbf{p}_j$. This will show that the set of extreme pathways is unique. For if $\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ was another set of extreme pathways the above argument would show that each of the $\mathbf{c}_i$ must be a positive multiple of one of the $\mathbf{p}_i$. This completes the proof of the theorem establishing the uniqueness of the extreme pathways.

APPENDIX B

The algorithm that is implemented to determine the set of extreme pathways for a reaction network follows the principles of algorithms for finding the extremal rays/generating vectors of convex polyhedral cones.

The algorithm begins with the formulation of an initial matrix consisting of an $n \times n$ identity matrix (**I**) appended to the transpose of the stoichiometric matrix, $\mathbf{S}^{\mathrm{T}}$. Then we examine the constraints on each of the exchange fluxes as given in eqn (4). If the exchange flux is constrained to be positive nothing is done; however, if the exchange flux is constrained to be negative then we multiply the corresponding row of the initial matrix by -1. If the exchange flux is unconstrained then we move the entire row to a temporary matrix, $\mathbf{T}^{(E)}$. This completes the initialization of the first tableau, $\mathbf{T}^{(0)}$. For the reaction system in Fig. 1, $\mathbf{T}^{(0)}$ and $\mathbf{T}^{(E)}$

are follows:

$$\mathbf{T}^{(0)} = \left[\begin{array}{cccccc|ccccc} 1 & & & & & & -1 & 1 & 0 & 0 & 0 \\ & 1 & & & & & 0 & -1 & 1 & 0 & 0 \\ & & 1 & & & & 0 & 1 & -1 & 0 & 0 \\ & & & 1 & & & 0 & 0 & -1 & 1 & 0 \\ & & & & 1 & & 0 & 0 & 1 & -1 & 0 \\ & & & & & 1 & 0 & 0 & -1 & 0 & 1 \end{array}\right], \tag{B.1}$$

$$\mathbf{T}^{(E)} = \left[\begin{array}{cccccc|ccccc} & & & 1 & & & -1 & 0 & 0 & 0 & 0 \\ & & & & 1 & & 0 & -1 & 0 & 0 & 0 \\ & & & & & 1 & 0 & 0 & -0 & -1 & 0 \\ & & & & & & 0 & 0 & -0 & 0 & -1 \end{array}\right]$$

We will designate each element of the above matrices by T_{ij}. Starting with x equal to 1 and $\mathbf{T}^{(0)}$ equaling $\mathbf{T}^{(x-1)}$ the next tableau is generated in the following manner:

1. Identify all of the metabolites that do not have an unconstrained exchange flux associated with them. The total number of such metabolites is denoted by μ. In this example, only metabolite C does not have such an unconstrained exchange flux so $\mu = 1$.

2. Begin forming the new matrix $\mathbf{T}^{(x)}$ by copying all rows from $\mathbf{T}^{(x-1)}$ which contain a zero in the column of $\mathbf{S}^{\mathrm{T}}$ that corresponds to the first metabolite identified in step 1, denoted by the index c. (This will be the third column of the transposed stoichiometric matrix.)

3. Of the remaining rows in $\mathbf{T}^{(x-1)}$ add together all possible combinations of rows which contain values of the opposite sign in column c, such that the addition produces a zero in this column. Given two rows, $\mathbf{r}_1$ and $\mathbf{r}_2$, whose elements will be denoted by $r_{1,j}$ and $r_{2,j}$ for $j = 1, \ldots, (n+m)$, combine the rows using the following equation to generate a new row $\mathbf{r}$ to be added to $\mathbf{T}^{(i)}$:

$$\mathbf{r}' = (|r_{2,c}| * \mathbf{r}_1) + (|r_{1,c}| * \mathbf{r}_2) \tag{B.2}$$

The resulting matrix $\mathbf{T}^{(1)}$ is as follows up to this point:

$$\mathbf{T}^{(0)} = \left[\begin{array}{cccccc|ccccc} 1 & & & & & & -1 & 1 & 0 & 0 & 0 \\ & 1 & 1 & & & & 0 & 0 & 0 & 0 & 0 \\ & 1 & & 1 & & & 0 & -1 & 0 & 1 & 0 \\ & 1 & & & & 1 & 0 & -1 & 0 & 1 & 1 \\ & & 1 & & 1 & & 0 & 1 & 0 & -1 & 0 \\ & & & 1 & 1 & & 0 & 0 & 0 & 0 & 0 \\ & & & & 1 & 1 & 0 & 0 & 0 & -1 & 1 \end{array}\right] \tag{B.3}$$

4. For all of the rows added to $\mathbf{T}^{(x)}$ in steps 2 and 3 check to make sure that no row exists that is a non-negative combination of any other sets of rows in $\mathbf{T}^{(x)}$. One method used is as follows: let $A(i)$ equal the set of column indices, j, for which the elements of row i equal zero. Then check to determine if there exists another row (h) for which $A(i)$ is a subset of $A(h)$. This is expressed mathematically in eqn (B.4) (analogous to the condition of eqns (14) and (15) in Schuster & Schuster, 1993). Thus, if eqn (B.4) holds true for any distinct rows i and h, then row i must be eliminated from $\mathbf{T}^{(i)}$

$$A(i) \subseteq A(h), \quad i \neq h$$

where

$$A(i) = \{ j: T_{i,j} = 0, \quad 1 \leqslant j \leqslant (n+m) \}. \quad \text{(B.4)}$$

5. With the formation of $\mathbf{T}^{(x)}$ complete repeat steps 2–4 for all of the metabolites that do not have an unconstrained exchange flux operating on the metabolite, incrementing x by one up to μ. The final tableau will be $\mathbf{T}^{(\mu)}$. [In this example there is only one such metabolite so we do not need to iterate through steps 2–4 again. Therefore $\mathbf{T}^{(\mu)}$ equals $\mathbf{T}^{(1)}$ as in eqn (B.3).)] Note that the number of rows in $\mathbf{T}^{(\mu)}$ will be equal to (k), the number of extreme pathways.

6. Next we append $\mathbf{T}^{(E)}$ to the bottom of $\mathbf{T}^{(\mu)}$ (also the same as $\mathbf{T}^{(1)}$ in this example). This results in the following tableau:

$$\mathbf{T}^{(l/E)} = \left[\begin{array}{ccccccc|ccccc}
1 & & & & & & & -1 & 1 & 0 & 0 & 0 \\
 & 1 & 1 & & & & & 0 & 0 & 0 & 0 & 0 \\
 & 1 & & 1 & & & & 0 & -1 & 0 & 1 & 0 \\
 & 1 & & & & 1 & & 0 & -1 & 0 & 0 & 1 \\
 & & 1 & & 1 & & & 0 & 1 & 0 & -1 & 0 \\
 & & & 1 & 1 & & & 0 & 0 & 0 & 0 & 0 \\
 & & & & 1 & 1 & & 0 & 0 & 0 & -1 & 1 \\
\hline
 & & & & & & 1 & -1 & 0 & 0 & 0 & 0 \\
 & & & & & 1 & & 0 & -1 & 0 & 0 & 0 \\
 & & & & 1 & & & 0 & 0 & 0 & -1 & 0 \\
 & & & 1 & & & & 0 & 0 & 0 & 0 & -1
\end{array}\right] \quad \text{(B.5)}$$

7. Starting in the $n+1$ column (or the first non-zero column on the right side), if $T_{i,(n+1)}$ does not equal to zero, then add the corresponding non-zero row from $\mathbf{T}^{(E)}$ to row i so as to produce a zero in the $(n+1)$ column. This is done by simply multiplying the corresponding row in $\mathbf{T}^{(E)}$ by $T_{i,(n+1)}$ and adding this row to row i. Repeat this procedure for each of the rows in the upper portion of the tableau so as to create zeros in the entire upper portion of the $(n+1)$ column. When finished, remove the row in $\mathbf{T}^{(E)}$ corresponding to the exchange flux for the metabolite just balanced.

8. Follow the same procedure as in step 7 for each of the columns on the right side of the tableau containing non-zero entries. (In this example we need to perform step 7 for every column except the middle column of the right side which corresponds to metabolite C.) The final tableau, $\mathbf{T}^{(Final)}$, will contain the transpose of the matrix $\mathbf{P}$ containing the extreme pathways in place of the original identity matrix. Both $\mathbf{T}^{(Final)}$ and $\mathbf{P}$ are given below:

$$\mathbf{T}^{(Final)} = \left[\begin{array}{cccccc|ccccc} 1 & & & & & -1 & 1 & & & 0 & 0 & 0 & 0 & 0 \\ & 1 & 1 & & & & & & & 0 & 0 & 0 & 0 & 0 \\ & 1 & & 1 & & & -1 & 1 & & 0 & 0 & 0 & 0 & 0 \\ & 1 & & & & 1 & -1 & & 1 & 0 & 0 & 0 & 0 & 0 \\ & & 1 & & 1 & & 1 & -1 & & 0 & 0 & 0 & 0 & 0 \\ & & & 1 & 1 & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & 1 & 1 & & -1 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{B.6}$$

$$\mathbf{P}^{\mathrm{T}} = \begin{array}{c} \begin{array}{cccccc|cccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & b_1 & b_2 & b_3 & b_4 & \end{array} \\ \left[\begin{array}{cccccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -1 & 1 \end{array}\right] \end{array} \begin{array}{l} \leftarrow \mathbf{p}_1 \\ \leftarrow \mathbf{p}_7 \\ \leftarrow \mathbf{p}_3 \\ \leftarrow \mathbf{p}_2 \\ \leftarrow \mathbf{p}_4 \\ \leftarrow \mathbf{p}_6 \\ \leftarrow \mathbf{p}_5 \end{array} \tag{B.7}$$

As a corollary to this algorithm it is easily seen that if an unconstrained exchange flux existed for every metabolite in the system the number of extreme pathways would simply equal to the number of internal fluxes in the system with each pathway equivalent to a single internal flux. The justification to ensure that this algorithm generates a unique set of systemically independent generating vectors for the flux cone is provided below. First we will show the existence of a conical generating set. Consider the cone

$$C = \{\mathbf{x} \in R^n : \mathbf{A}\mathbf{x} = \mathbf{0},\ \mathbf{x} \geqslant \mathbf{0}\}, \tag{B.8}$$

where $\mathbf{A}$ is an $m \times n$ matrix. We will induct on the number of rows of $\mathbf{A}$. If $m = 0$ then C is just the positive orthant, and the set of coordinate vectors is easily shown to be a generating set for C.

For the induction step we have a cone C with generating set $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$. Let

$$C' = C \cap \{\mathbf{x} : a_1 x_1 + \cdots + a_n x_n = \mathbf{a} \cdot \mathbf{x} = 0\}. \tag{B.9}$$

After a possible rescaling of the $\mathbf{p}_i$ and recording, we can assume that

$$\mathbf{a} \cdot \mathbf{p}_i = \begin{cases} 1, & 1 \leqslant i \leqslant k, \\ -1 & k+1 \leqslant i \leqslant k+l, \\ 0 & k+l+1 \leqslant i \leqslant n. \end{cases} \tag{B.10}$$

It is easily shown that the pathways

$$\{\mathbf{p}_i + \mathbf{p}_j : 1 \leqslant i \leqslant k,\ k+1 \leqslant j \leqslant k+l\} \cup \{\mathbf{p}_i : k+l+1 \leqslant i \leqslant n\} \tag{B.11}$$

form a conical-generating set for the cone C'. Using the induction step for each row of **A**, we construct a conical-generating set. This set will most likely not be systemically independent. However, we can remove pathways from the set to get a subset that still generates the cone and is systemically independent. This procedure will produce a systemically independent generating set for the cone.

APPENDIX C: NOTATION

$\mathbf{0}$	zero vector
b_j	activity of the jth exchange flux
C	steady-state flux cone
d	dimension of the null space of the stoichiometric matrix
k	number of extreme pathways
m	number of reaction metabolites
n	total number of fluxes
n_E	number of exchange fluxes
n_I	number of internal fluxes
$\mathbf{P}$	matrix of extreme pathway
$\mathbf{p}_i$	the ith extreme pathway
r	rank of the stoichiometric matrix
$\mathbf{S}$	stoichiometric matrix
S_{ij}	element of the ith row and the jth column of the stoichiometric matrix
$\mathbf{v}$	flux vector
v_j	activity of the jth internal flux
$\mathbf{w}$	pathway utilization vector
w_i	activity of the ith pathway
$\mathbf{x}$	concentration vector.